

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 December 2002 (12.12.2002)

PCT

(10) International Publication Number
WO 02/099398 A1

(51) International Patent Classification⁷: **G01N 21/64**
(21) International Application Number: PCT/US02/17124
(22) International Filing Date: 31 May 2002 (31.05.2002)
(25) Filing Language: English
(26) Publication Language: English

(30) Priority Data:
09/875,779 6 June 2001 (06.06.2001) US

(71) Applicant: U.S. GENOMICS, INC. [US/US]; 6H Gill Street, Woburn, MA 01801 (US).

(72) Inventors: CHAN, Eugene, Y.; 133 Park Street, #1001, Brookline, MA 02446 (US). GLEICH, Lance, C.; 90 College Avenue #2, Somerville, MA 02144 (US). WELLMAN, Parris, S.; 61-3A Taurus Drive, Hillsborough, NJ 08876 (US).

(74) Agent: LOCKHART, Helen, C.; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

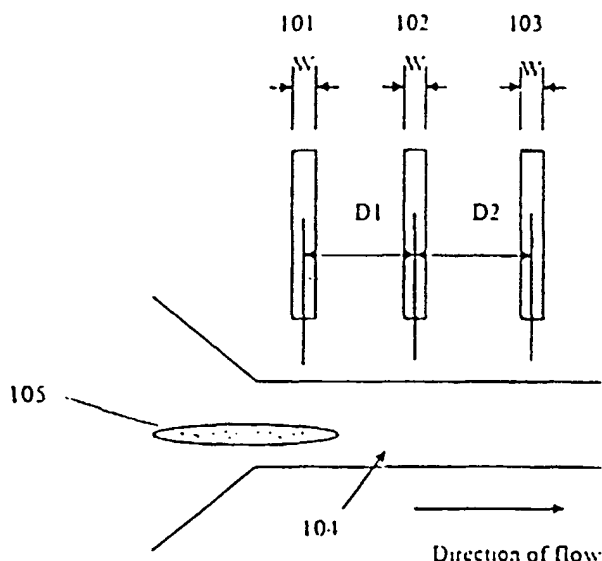
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND APPARATUS FOR CHARACTERIZATION OF SINGLE POLYMERS



(57) Abstract: In a method and apparatus for determining the velocity of an elongated polymeric molecule, the molecule (105) is moved relative to one or more detection stations (101, 102, 103). Time correlated detection signals are used to determine velocity, length, and distance between landmarks on the molecule (105).

WO 02/099398 A1

**METHODS AND APPARATUS FOR CHARACTERIZATION
OF SINGLE POLYMERS**

FIELD OF THE INVENTION

5 The present invention relates generally to the field of molecular analysis and characterization. More particularly, the present invention relates to methods and apparatus for determining the velocity of elongated polymeric molecules moving relative to a detection station, as well as to methods and apparatus for determining the length of such molecules and the distance between landmarks present on such molecules.

10

BACKGROUND OF THE INVENTION

Analysis of the structure and dynamics of single macromolecules in a fluid sample has attracted considerable interest due in part to the rapid development of methodologies for the manipulation and detection of single macromolecules. For example, recent developments
15 in experimental techniques and available hardware have increased dramatically the sensitivity of detection so that optical detection can be made of single dye molecules in a sample. Single dye detection can be done in an aqueous solution, at room temperature (see, e.g., Weiss, 1999, Science 283: 1676-1683), and in very small volumes to reduce background. Such single-molecule based analytical methods are especially useful in the analysis of biological
20 macromolecules, such as nucleic acid molecules and proteins. Single-molecule analytical methods require small amounts of sample, thereby alleviating tedious efforts in generating large amounts of sample material. For example, single-molecule analytical methods may allow analysis of the structure of nucleic acid molecules without amplification, e.g., by polymerase-chain reaction (PCR). Single-molecule analytical methods also allow analysis of
25 individual molecules, and are thus particularly useful in the identification of structural and/or dynamical features without the effect of averaging over a heterogeneous population.

A single-molecule electrophoresis (SME) method which combines single molecule detection and electrophoresis has been reported for the detection and identification of single molecules in solution (Castro and Shera, 1995, Anal. Chem. 67: 3181-3186). In SME, sizing
30 of single molecules is accomplished through determination of electrophoretic velocities by measuring the time required for individual molecules to travel a fixed distance between two laser beams. This method has been applied to DNA, to fluorescent proteins and to simple

organic fluorophores. For example, SME offers a single-molecule method for sizing of DNA restriction fragments. However, SME detects only the presence or absence of a molecule. The method does not provide information regarding the internal structure of a molecule.

5 A single-molecule DNA sizing method using a microfabricated device has also been reported (Chou et al., 1999, Proc. Natl. Acad. Sci. USA 96:11-13). The method makes use of the fact that the amount of intercalated dye is proportional to the length of the molecule, and determines the lengths of single DNA molecules by measuring the total fluorescence intensity of DNA stained with intercalating dye molecules. Thus, the method does not use electrophoretic mobilities to determine sizes of molecules. This method also does not
10 provide information regarding the internal structure of a molecule.

PCT Publication No. WO 98/10097 discloses a method and apparatus for detection of single molecules emitting two-color fluorescence and determination of molecular weight and concentration of the molecules. The method involves labeling of individual molecules with at least two fluorescent probes of different emission spectra. Simultaneous detection of the
15 two labels indicates the presence of the molecule. The velocity of the molecule is determined by measuring the time required for the molecules to travel a fixed distance between two laser beams. Comparison of the molecule's velocity with that of standard species permits determination of the molecular weight of the molecule, which may be present in a concentration as small as one femtomolar.

20 Other techniques for characterizing single macromolecules include a method described in U.S. Pat. No. 5,807,677 for direct identification of a specific target nucleic acid sequence having a low copy number in a test solution. This method involves the preparation of a reference solution of a mixture of different short oligonucleotides. Each oligonucleotide includes a sequence complementary to a section of the target sequence and is labeled with
25 one or more fluorescent dye molecules. The reference solution is incubated with the test solution under conditions favorable to hybridization of the short oligonucleotides with the nucleic acid target. The target sequence is identified in the solution by detection of the nucleic acid strands to which one or more of the labeled oligonucleotides are hybridized. To amplify the fluorescence signal, a "cocktail" of different oligonucleotides is used. In this
30 cocktail, the oligonucleotides are capable of hybridizing with sequences adjacent to but not overlapping with the target sequence. The disadvantage of this method is that, in order to design probes of the proper sequence, the exact sequence of the target nucleic acid and surrounding sequences must be known. A method described in U.S. Pat. No. 5,599,664 and

European Patent No. EP 0391674 allows sizing of DNA molecules by first subjecting a DNA molecule to a force such that the DNA molecule is elongated and then measuring the conformational relaxation dynamics. In another method (Schmalzing et al., 1998, Analytical Chemistry 70:2303-2310; Schmalzing et al, 1997, Proc. Natl. Acad. Sci. USA 94:10273-10278), microfabricated devices for DNA analysis were developed, including sequencing, which employ small-scale versions of traditional techniques, such as electrophoresis.

None of these single molecule analytical methods allows the determination of the internal structure of the molecule. A challenge to the characterization of the internal structure, e.g., the linear sequence of monomers, in a single polymer chain is the natural tendency of polymers in most media to adopt coiled conformations. The average degree of such coiling is dependent on, inter alia, the interaction of the polymer with the surrounding solution, the rigidity of the polymer, and the energy of interaction of the polymer with itself. In most cases, the coiling is quite significant. For example, a λ -phage DNA, with a B-form contour length of about 16 μ m long, has a random coil diameter of approximately 1 μ m in water (Smith et al., 1989, Science 243:203-206).

Methods of elongating DNA molecules by fluid flow have been reported (Perkins et al. Science 276:2016-2021; Smith et al., Science 283:1724-1727). In one method, DNA molecules are stretched by an elongational flow. The probability distribution of molecular extension was determined as a function of time and strain rate. Detailed dynamics of elongated DNA molecules in elongational flow has also been observed. In another method DNA molecules are stretched by a steady shear flow. The probability distribution for the molecular extension was determined as a function of shear rate. It was found that, in contrast to the behavior in pure elongational flow, the average polymer extension in shear flow does not display a sharp coil-stretch transition.

DNA has also been stretched by electrophoresis as part of a near-field detection scheme for sequencing biomolecules. DNA has been elongated by electrophoresis both in a gel and in solution, using electrical forces to move the DNA in position for reading (U.S. Pat. No. 5,538,898). However, no data were given to determine the quality of the stretching of large polymers, and the technique is limited to analyzing approximately 3 megabases at a time.

Gravitational forces have also been used to stretch DNA (U.S. Pat. No. 5,707,797; Windle (1993) Nature Genetics 5:17-21). In this technique, drops of DNA from the sodium dodecyl sulfate lysing of cells were allowed to run down a slide held at an angle. The effect

of gravity was enough to stretch out the DNA, even to its over-stretched S-DNA form. The DNA was then immobilized on the slide, making processing, e.g., fluorescent labeling, prior to stretching relatively difficult.

Single-molecule DNA analytical methods which involve elongation of DNA molecule include optical mapping (Schwartz et al., 1993, Science 262:110-113; Meng et al., 1995, Nature Genet. 9:432; Jing et al., Proc. Natl. Acad. Sci. USA 95:8046-8051) and fiber-fluorescence in situ hybridization (fiber-FISH) (Bensimon et al., Science 265:2096; Michalet et al., 1997, Science 277:1518). In optical mapping, DNA molecules are elongated in a fluid sample and fixed in the elongated conformation in a gel or on a surface. Restriction digestions are then performed on the elongated and fixed DNA molecules. Ordered restriction maps are then generated by determining the size of the restriction fragments. In fiber-FISH, DNA molecules are elongated and fixed on a surface by molecular combing. Hybridization with fluorescently labeled probe sequences allows determination of sequence landmarks on the DNA molecules. Both methods require fixation of elongated molecules so that molecular lengths and/or distances between markers can be measured.

A method for measuring the length and distances between markers on DNA was developed by Kambara et al. (U.S. Pat. No. 5,356,776). This method involves fluorescently labeling a DNA molecule at both termini and/or internal sites, and moving the labeled molecule through a gel via electrophoresis. In so doing, the DNA molecule is forced into a straightened conformation. The straightened DNA molecule is transferred into a gel-free buffer, and the fluorescent labels are detected. The time interval between the detection of the two labels is used to determine the distance between them. If the two labels label the termini of the DNA molecule, the distance between the labels measures the length of the molecule. The method, which does not provide means for determining the velocity of the DNA molecule, relies on estimating the velocity of DNA from the migration rate of the DNA molecule.

Flow based single-molecule analytical methods for elongation and characterization of single macromolecules have not been widely adopted due in part to the difficulty in precise measurement of molecular characteristics, e.g., the length of the macromolecule, the distance between two landmarks on a macromolecule, etc. For example, to determine the length of an elongated macromolecule as it travels through a detection zone, e.g., a laser excitation zone, it is necessary to know the velocity of the macromolecule. The flow velocity field can be measured by various known methods, e.g., particle image velocimetry (PIV) (see, e.g.,

Meinhart et al., 1999, Experiments in Fluids 27: 414-419; Meinhart et al., 2000, Meas. Sci. Technol. 11:809-814). The velocities of flexible objects, such as elongated polymers, may not be the same as the flow velocities. For example, in most flows the length of a polymer may be changing as it travels along with flow. In particular, the length of a polymer may be changing as a consequence of changing flow velocity. There is therefore a need for faster, simpler, more reliable and more universally applicable methods for measuring the velocities of single elongated polymers traveling in a flow. There is also a need for more accurate methods for determining the length of single elongated polymers and/or distances between landmarks on single elongated polymers.

SUMMARY OF THE INVENTION

The present invention relates to methods and apparatus for the characterization of single polymers. In particular, the invention relates to methods and apparatus for determination of the velocities of single elongated polymers as they are caused to move relative to detection stations. The invention also relates to methods and apparatus for determining the length and molecular mass of single polymers, as well as to methods and apparatus for determining the distance between landmarks present on such polymers.

Broadly, the velocity of an elongated polymer, such as a DNA molecule having at least one detectable label, can be determined by a method in which the molecule is moved relative to at least two linearly sequential detection zones that are spaced apart a predetermined distance. As the label moves relative to a detection zone, it interacts with the zone to produce a signal amplitude profile. Each such profile is characterized in that it includes a leading edge in which the signal is increasing during a first time interval, a detection amplitude in which the signal remains substantially constant during a second time interval, and a trailing edge in which the signal is decreasing during a third time interval. A representation of the signal amplitude over the course of the first, second and third time intervals defines the signal amplitude profile. Each of the signal amplitude profiles is measured in a time-correlated manner, and then the time-correlated measurements can be analyzed to determine the velocity of the polymer.

Thus, in one embodiment, the signal amplitude profile from an individual label is determined at two linearly sequential detection zones that are spaced apart a predetermined distance. By measuring the time difference for a selected element of the signal amplitude profile to be detected at a first detection zone and at a second detection zone, and by dividing

the predetermined distance between the zones by the measured time, the velocity of the polymer is determined. The selected element of the signal amplitude profile may be the center-of-mass of the profile, the temporal center of the profile, or some point, i.e., the half-way point, during the signal increase as the leading edge of the profile is detected or the
5 signal decrease as the trailing edge of the profile is detected. It should be noted that the label need not be limited to a particular location. By comparing the signal amplitude profile characteristics at two different detection zones, the label can be such that it is detected for only a brief time, as for example, a single labeled polymeric base is detected, or it may be detected for a longer period of time, as for example, a polymer labeled with an intercalating
10 dye travels entirely through two detection zones.

In a second embodiment of the present invention, the length of the polymer can be determined. In this embodiment, along with detecting an individual label on the polymer at two detection zones, the polymer must be detectable whenever any portion of it is positioned relative to a detection zone. Thus, for example, the polymer may be labeled with an
15 intercalating dye that is detectable along the entire length of the polymer molecule. As the molecule moves relative to the detectors, the time difference between the detection of a selected element of the signal amplitude profile at the first detection zone and at a second detection zone is measured, and by dividing the predetermined distance between the zones by the measured time, the velocity of the polymer is determined. This value is then multiplied
20 by the total time during which the polymer is detected at either of the detection zones, and the resulting product is the length of the molecule.

In a third embodiment of the present invention, the length between two detectable landmarks on the polymer can be determined. In this embodiment, two detectable landmarks are positioned on the polymer, and then the polymer is moved relative to two linearly
25 sequential detection zones that are spaced apart by a predetermined distance. The polymer can be labeled with an intercalating dye or the like along its entire length, however, this is optional. The velocity of the polymer is determined using the methods described in the first embodiment above (using signal amplitude profiles obtained from either one of the landmarks or the optional intercalating dye), and the time difference between detection of one
30 of the landmarks at a first detector and detection of that same landmark at a second detector is measured. When the calculated velocity is multiplied by the measured time difference between landmark detection, the resulting product is the distance between the landmarks.

Each of the limitations of the invention can encompass various embodiments of the invention. It is, therefore, anticipated that each of the limitations of the invention involving any one element or combinations of elements can be included in each aspect of the invention.

5 BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 is a schematic representation of the apparatus of the invention.

FIG. 2 is a schematic representation of several signal amplitude profiles measured using the apparatus of FIG. 1.

FIG. 3 is a schematic illustration of several time-correlated signal amplitude profiles.

FIG. 4 is a schematic illustration of several time-correlated signal amplitude profiles measured at multiple detection zones.

FIG. 5 is a schematic illustration of signals detected at first and second detection zones.

FIG. 6 is a schematic illustration of one embodiment of an apparatus for formation of
15 laser spots at multiple detection zones.

FIG. 7 is a schematic illustration of a second embodiment of an apparatus for formation of laser spots at multiple detection zones.

FIG. 8 is a schematic illustration of a third embodiment of an apparatus for formation of laser spots at multiple detection zones.

FIG. 9 is a schematic illustration of one embodiment of an apparatus to generate interference fringes.

FIG. 10 is a schematic representation of the apparatus of the invention showing positioning of laser spots defining the detection zones.

FIGS. 11A-11C show exemplary signal amplitude profiles measured at two detection
25 zones.

FIG. 12 shows a velocity histogram based upon center-of-mass velocity of a population of λ -DNA molecules measured using one method of the invention.

FIG. 13 shows a length histogram of a population of λ -DNA molecules measured using one method of the invention.

30

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides methods and apparatus for determining the velocities of single elongated polymeric macromolecules. The methods of the invention are based on

time-correlated measurements of an elongated polymeric macromolecule at each of a plurality of detection zones. The detection zones are located along the travel path of the polymer at predetermined spacings. Signal amplitude profiles, e.g., intensity-time curves when fluorescence-based measurements are used, of an elongated polymeric macromolecule, are measured as the macromolecule passes through each of the detection zones. The measurements in the plurality of detection zones are time-correlated, e.g., synchronized, so that the temporal spacings between signal amplitude profiles measured at different detections zones are also determined. A schematic illustration of the invention is presented in FIG. 1. This apparatus is used to generate signal amplitude profiles as shown in FIG. 2. In FIG. 2, the signal amplitude profiles are positive, however, it should be apparent to one skilled in the art that negative signal amplitude profiles are also encompassed. For example, when absorption-based measurements are used, the signal amplitude will decrease when the polymer is in the detection zone.

As used herein, the term "elongated polymeric macromolecule" or "stretched polymeric macromolecule" refers to a polymer that is in a conformation in which the length of the molecule is substantially greater than its radius of gyration in the same solvent. Preferably the elongated polymeric macromolecule has a length which is substantially close to its contour length. An elongated polymer may be passed through an elongated channel. An "elongated channel" or "elongation channel" is a channel or pore through which a polymer can traverse. The channel, in some embodiments, has an opening which is only slightly larger than the cross-section of the elongated polymer.

As used herein, a "detection zone" refers to a region or volume which experiences movement relative to the polymeric macromolecule. Generally, the detection zones are fixed along a traveling path of a macromolecule, however, the invention contemplates a fixed macromolecule and traveling detection zones. The detection zone acts as a region in which a signal from the macromolecule is measured if the macromolecule, or a portion thereof, is present in the region or volume. In embodiments based on measuring laser induced fluorescence, a detection zone is defined by the excitation volume of a laser beam and the signal is the total fluorescence intensity emitted by the macromolecule or a portion thereof as it passes through the excitation volume.

As used herein, "time-correlated measurements" refers to measurements performed in different detection zones in such a manner that the temporal spacings between measured signals in different detections zones are known. Such time-correlated detection can be

achieved by synchronizing measurements in all detection zones in real time. Alternatively, measurements in different detection zones may be time-correlated by a fixed, known time delay of detection in different detection zones. Methods for synchronized and time-delayed detection in different detection zones are well known in art and will be apparent to one skilled
5 in art.

As used herein, a "signal amplitude profile" refers to the temporal profile of a measured signal. Each such profile is characterized in that it includes a leading edge in which the signal is increasing during a first time interval, a detection amplitude in which the signal remains substantially constant during a second time interval, and a trailing edge in
10 which the signal is decreasing during a third time interval. A representation of the signal amplitude over the course of the first, second and third time intervals defines the signal amplitude profile. In embodiments involving optical detection, e.g., when macromolecules are detected by fluorescence, the signal amplitude profile is represented by the intensity versus time curve measured by a photodetector. Preferably, contributions to a signal
15 amplitude profile due to the finite size of a detection zone are eliminated. In embodiments involving laser excitation, the laser beam profile is preferably deconvoluted from a measured intensity-time curve. Methods for eliminating the contribution of the finite size of a detection zone to a signal amplitude profile, e.g., a laser beam profile, are well known in art and can be performed by one skilled in the art.

20 The present invention also provides methods and structures that allow polymers of any length, including nucleic acids containing entire genomes, to undergo measurement of their velocities and lengths.

In carrying out the method of the invention, polymers are loaded into a device and run through channelled structures, propelled by forces which may include, but not be limited to
25 kinetic, electrical or chemical forces. Stretching or elongation of the polymer can be achieved using any of the methods described in co-pending application USSN 09/636,793. The stretching forces on the polymer are such that, in the case of an elongated DNA molecule, the molecule may be stretched to lengths much longer than the dimension of the detection zone. Since multiple molecules may be stretched in succession, extremely high
30 throughput screening, e.g., screening of more than one molecule per second, can be achieved.

In the invention, single extended, labeled polymers are moved past detection zones. At such zones, labeled units of the polymers interact with the detector to produce a signal. As used in this application, "moved past" refers to embodiments in which the detection zones

are stationary and the extended polymers are in motion, as well as to embodiments in which the detection zones are in motion and the extended polymers are stationary. Embodiments in which the detection zones and extended polymers are both in motion are contemplated as well.

5 Although the invention may be used for characterizing any polymer, it is preferable that the polymers have a predominantly, though not necessarily exclusively, linear or single-chain arrangement. Examples of such polymers include biological polymers such as deoxyribonucleic acids, ribonucleic acids, polypeptides, and oligosaccharides. The polymers may be heterogeneous in backbone composition, thereby containing any possible
10 combination of individual monomer units linked together, e.g., peptide-nucleic acids (PNA), which have amino acids linked to nucleic acids. In a preferred embodiment, the polymers are homogeneous in backbone composition and are, e.g., nucleic acids, polypeptides or oligosaccharides. The term "backbone" is given its usual meaning in the field of polymer chemistry. A nucleic acid as used herein is a biopolymer comprised of nucleotides, such as
15 deoxyribose nucleic acid (DNA) or ribose nucleic acid (RNA). A protein or polypeptide as used herein is a biopolymer comprised of amino acids. In the most preferred embodiment, the extended object is a double-stranded DNA molecule.

 A polymer is made up of a plurality of individual units, i.e., monomeric units or monomers, which are building blocks that are linked either directly or indirectly to other
20 building blocks or monomers to form the polymer. The polymer preferably comprises at least two chemically distinct linked monomers. The at least two chemically distinct linked monomers may produce or be labeled to produce different signals. Different types of polymers are composed of different monomers. For example, DNA is a biopolymer comprising a deoxyribose phosphate backbone to which are attached purines and pyrimidines
25 such as adenine, cytosine, guanine, thymine, 5-methylcytosine, 2-aminopurine, hypoxanthine, and other naturally and non-naturally occurring nucleobases, substituted and unsubstituted aromatic moieties. RNA is a biopolymer comprising a ribose phosphate backbone to which are attached purines and pyrimidines such as those described for DNA but wherein uracil is substituted for thymidine. Deoxyribonucleotides may be joined to one another via an ester
30 linkage through the 5' or 3' hydroxyl groups to form the DNA polymer. Ribonucleotides may be joined to one another via an ester linkage through the 5', 3' or 2' hydroxyl groups. Alternatively, DNA or RNA units having a 5', 3' or 2' amino group may be joined via an amide linkage to other units of the polymer.

The polymers may be naturally-occurring or non-naturally occurring polymers. Polymers can be isolated, e.g., from natural sources using biochemical purification techniques. Alternatively, polymers may be synthesized, e.g., enzymatically by in vitro amplification using the polymerase chain reaction (PCR), by chemical synthesis, or by
5 recombinant techniques.

The structures of the invention are used in conjunction with methods for analyzing the extended polymers by detecting a physical quantity which transmits or conveys information about the structural characteristics of an extended polymer. A physical quantity, as used herein, can either be a measurable intrinsic property of a particular type associated with one
10 or more monomers of an extended polymer, e.g., the distinct absorption maxima of the naturally occurring nucleobases of DNA (referred to herein as intrinsic labeling), or a measurable property of a compound that is specifically associated with one or more monomers of an extended polymer (referred to herein as extrinsic labeling). Preferably the physical quantity is proportional to the number of monomers in the detection zone. An
15 extrinsically labeled polymer may be labeled with a particular fluorescent dye with which all nucleobases of a particular type, e.g., all thymine nucleobases, in a DNA strand are labeled. Alternatively, an extrinsically labeled polymer may be a fluorescently labeled oligonucleotide of defined length and sequence that hybridizes to and therefore "marks" the complementary sequence present in a target DNA. An extrinsic label of this sort is referred to as landmark or
20 a unit specific marker. Thus, a landmark is an extrinsic label that identifies a specific sequence within the polymer that is more than one unit in length. Molecules, or markers, that can be used to labeled polymers may further include, but are not limited to, sequence specific major or minor groove binders and intercalators, sequence-specific DNA or peptide binding proteins, sequence specific PNAs, etc. The detectable physical quantity may be in any form
25 that is capable of being measured. For instance, the detectable physical quantity may be electromagnetic radiation, electrical conductance, heat conductance, radioactivity, etc. The measured signal may arise from energy transfer, directed excitation, quenching, changes in conductance (resistance), or any other physical changes. In one embodiment, the measured signal arises from fluorescence resonance energy transfer ("FRET") between the marker and
30 the station, or the environment surrounding the station. In preferred embodiments, the measured signal results from direct excitation in a confined or localized region, or epiillumination of a confocal volume or a slit-based excitation. Possible analyses of polymers include, but are not limited to: determination of polymer length, determination of

polymer sequence, determination of polymer velocity, determination of the degree of identity of two polymers, determination of characteristic patterns of unit-specific markers of a polymer to produce a "fingerprint", and characterization of a heterogeneous population of polymers using a statistical distribution of unit-specific markers within a sample population.

- 5 The exemplary labels include but are not limited to intercalator dyes: YOYO-1, TOTO-1, YOYO-3, TOTO-3, POPO-1, POPO-3, JOJO-1, JOJO-3, BOBO-1, BOBO-3 (from molecular probes); non-intercalator/backbone staining dyes (these dyes chemically attach to the backbone of DNA): Panvera-fluorescein kit, Panvera-Cy3 kit, Panvera-Cy 5 kit; fluorescent probe dyes: fluorescein, tetramethylrhodamine, Alexa dyes, Oregon dyes, Cy dyes, IR dyes; 10 and other types of labels, e.g., latex spheres, gold particles, and streptavidin-biotin conjugates.

There are numerous methods and products available for analyzing polymers as described in PCT Publication No. WO 98/35012.

- Various methods for analyzing polymers differ in their potential sensitivity and 15 resolution, e.g., the minimum distance between two detection zones. A low resolution technique is capable of measurements in two detection zones having a large distance between them; a high resolution technique is capable of measurements in two detection zones having a smaller distance between them. The resolution of a particular technique is determined by the characteristic distance through which the detection method may sense the particular physical 20 quantity of an extended polymer. For example, the resolution of optical methods is dictated by the diffraction limit.

In the following, for simplicity reasons, optical detection methods are often used. It will be apparent to one skilled in the art that other detection methods can be used in conjunction or in place of optical detections.

25

Velocities of Single Elongated Polymers

- The methods of the invention are based on time-correlated measurements of elongated polymers as they travel past two or more detection zones which are spaced apart by known, pre-determined distances. A schematic illustration of one embodiment of the apparatus of the 30 invention is presented in FIG. 1. For example, detection zones 101 and 102 are separated by a distance represented as D1. The corresponding time-correlated signal amplitude profiles 106 and 107 shown in FIG. 2 are temporally separated by the time the polymer (105) takes to travel the distance D1 between the detection zones. As the polymer passes each detection

zone, a signal amplitude profile is generated. These are depicted schematically in FIG. 2, in which the polymer 105 first passes detection zone 101 which generates profile 106, then passes detection zone 102 which generates profile 107, and finally passes detection zone 102 which generates profile 108. Signal amplitude profiles are then used to determine the
5 velocity of the polymer. Of course, the invention is not intended to be limited to the use of three detection zones. Rather, as few as two detection zones may be employed, while the maximum number of such zones is virtually unlimited.

As will be described in greater detail below, FIGS. 3, 4 and 5 illustrate various ways different types of velocity can be determined from measured signal amplitude profiles. The
10 distance between each pair of detection zones can be selected based on the expected test conditions and/or requirements of the measurements. For example, when a steady flow is involved such that the polymer velocity of polymers is not expected to vary over a range of distances, larger distance between the detection zones may be used. Alternatively, when flow changes rapidly and/or the polymers vary in length rapidly, multiple closely spaced detection
15 zones are preferred (as shown in FIG. 4). In a preferred embodiment, the distances between the detectors are less than the length of an elongated polymer. Shorter distances between the detectors tend to cause the average measured velocity to be closer to the actual measured velocity at both detectors. Thus in some preferred embodiments, the distances between detectors approach the minimum allowed by the resolution of the detection zones and their
20 detection methods. For example, in methods based on optical measurement, the distances between detection zones are at or close to the diffraction limit.

Distances among detection zones can be determined using any methods known in the art. In one embodiment, the distances are measured by a mechanical systems that offer high spatial resolution. In another embodiment, the distances can be measured by calibration
25 using elongated polymers of a known length. In still another embodiment, the distances can be measured by calibration using a fluid flow of known velocity. Such calibration methods are well known in the art.

Due to its flexibility, the conformation of a polymer may be constantly changing. As a consequence, each point on the polymer may have a different velocity from another point
30 on the polymer. For polymers in a fluid flow, the velocity of an elongated polymer is also affected by the fluid flow. The velocity of a single elongated polymer can be described in various manners.

In FIG. 3, various velocity measurements are represented for first 201 and second 202 signal amplitude profiles measured at first and second detection zones. Various time intervals, use for the velocity determination, are depicted in FIG. 3 as well.

5 The velocity of an elongated polymer can be represented by the velocity of its center-of-mass (COM). In one embodiment, the COM can be found by obtaining the signal amplitude profile as the molecule moves past each detector, and then calculating the time at which half of the sum of the signal intensities has occurred. The COM velocity can be measured by determining the time interval the center-of-mass of the elongated polymer requires to travel between two detection zones separated by a known distance. This time
10 interval is represented by the bar labeled reference numeral 204 in FIG. 3. The COM velocity is especially useful when the molecule elongates or shortens during passage through the system. For example, the COM velocity based on measurements of intercalator signal takes into account any variations in stretching because varying regions of stretching will have proportionately higher or lower intensity of the intercalator signal.

15 The velocity of an elongated polymer can also be represented by the velocity of its contour center, i.e., the midpoint of its molecular contour (center-to-center velocity). The center-to-center velocity can be measured by determining the time interval the contour center of an elongated polymer requires to travel between two detection zones separated by a known distance. This time interval is represented by the bar labeled reference numeral 203 in FIG.
20 3. Depending on whether the polymer is uniformly elongated or not, the center-to-center velocity can be the same or different from the COM velocity. In particular, if the polymer is not uniformly elongated, the center-to-center velocity is generally different from the COM velocity.

The velocity of an elongated polymer can also be represented by the velocity of its
25 leading or trailing end, i.e., end-to-end velocity. The end-to-end velocity can be measured by determining the time interval the leading or trailing end of an elongated polymer requires to travel between two detection zones separated by a known distance. This time interval is represented by the bar labeled reference numeral 206 in FIG. 3.

The velocity of an elongated polymer can also be represented by the rise-time velocity
30 which is defined as the velocity of the leading end of the elongated polymer traverse through a region of finite size. A rise-time velocity can therefore be measured by dividing the size of the region with the time required for the leading end of the polymer to travel from the entry edge of the region to the exit edge of the region. This time interval is represented by the bar

labeled reference numeral 205 in FIG. 3. Rise-time velocity is particularly useful when the leading end of the polymer is labeled.

Any single type of velocity can be used either independently or in combination with one or more other types of velocity for the characterization of an elongated polymer. The measured velocity can be used to determine the polymer length and/or distances between markers on an elongated polymer.

The methods of detection can be based on optical methods, including but not limited, to light induced fluorescence measurements, absorption measurements, light scattering measurements and non-linear optical measurements. Such optical methods are well known in the art. For example, when light induced fluorescence measurements or absorption measurements are to be used, those of ordinary skill in the art will know how to choose, the excitation source, the wavelength, the detector, and so on.

In some embodiments, the signal measured at the detection zone is the light induced fluorescence of an elongated polymer. In one embodiment, laser induced fluorescence of monomers in the polymer is measured. In another embodiment, laser induced fluorescence of intercalating dye molecules is measured. In still another embodiment, light induced fluorescence from fluorescent markers labeled at discrete locations along the polymer, e.g., at the leading or trailing ends, is measured. In other embodiments, a slit of selected width is used to define the detection zone. In this embodiment, the excitation can be selected from the emitted wavelengths of a lamp and the detection zone is illuminated through the slit. When a slit is used, it is preferred that the slit be of a width that is close to the diffraction limit of the wavelength of light used so that the highest resolution can be achieved.

In other embodiments, the signal measured at the detection zone can be based upon the absorption of an elongated polymer. In such embodiments, the measured wavelength or wavelengths of the absorbed light are proportional to the number of monomers falling within the detection region. In one embodiment, laser excitation is used in the absorption measurements. In this embodiment, a laser beam of a selected wavelength is used to define the detection zone. In another embodiment, a slit of select width is used to define the detection zone. As such, the excitation can be selected from the emitted wavelengths of a lamp and the detection zone is illuminated through the slit.

The signal measured at the detection zone can also be light scattered by a polymer labeled with molecules having light-scattering properties. As noted above, a laser beam of a

selected wavelength can be used to define the detection zone or, alternately, a slit of select width is used to define the detection zone.

Various methods can be used to define multiple detection zones in the present invention. In methods based on optical measurements, detection zones can be generated using laser illumination spots. In embodiments which employ multiple lasers, a separate laser can be used to generate each spot. In that embodiment, the lasers are arranged such that the collimated output of the lasers converges on the output aperture of a microscope objective. As such, each beam is caused to enter the microscope objective at a different angle. The angular displacement of the beams is converted by the objective lens to a spatial separation of the focused laser spots as shown schematically in FIG. 6. Specifically, in FIG. 5, two separate laser beams 300, 302 pass through a microscope objective lens 304 which resolves them into two separate spots 306, 308 on a sample plane 310. The spatial separation of the spots can be controlled by adjusting the angular separation of the beams. The larger the angular separation, the larger the spatial separation.

In other embodiments, multiple detection zones can be defined by splitting the output of a single laser to create angularly displaced beams. In one embodiment depicted schematically in FIG. 7, a first partially reflecting beamsplitter 320 and a fully reflective mirror 322 can be employed to create two beams 324, 326 that converge on a second partially reflecting beamsplitter 328 which reflects the beams onto the output aperture of a microscope objective 330. The angular separation is controlled by setting the spacing and angles between the first partially reflecting beamsplitter 320 and the fully reflecting mirror 322. In a preferred embodiment, the partially reflecting beamsplitter is a pellicle type of beamsplitter. The very thin membrane of the pellicle means that a ghost beam from the rear surface of the beamsplitter is effectively avoided.

In still another embodiment, such as that shown schematically in FIG. 8, a diffractive optical element such as a phase grating is used to split a monochromatic laser beam into multiple beams at various grating orders, e.g., a zero order (straight through) beam, first order beams and higher order beams. The distribution of energy into the different orders is determined by the depth of grooves formed in the phase grating, whereas the angular separation of different orders of beams is determined by groove spacing. In this embodiment a collimated laser beam is bent by a fully reflective mirror 330 and passed through a phase grating 332. A second optical element, such as a lens 334, is used to collect the diverging beams from a diffractive optics and refocus the beams 336, 338, so that they converge

encounter a partially reflecting beamsplitter 340 and are reflected toward a microscope objective 340.

In yet another embodiment, shown schematically in FIG. 9, a fringe pattern of alternating dark and bright bands formed within a single illumination spot 350 is used to define detection zones. An elongated polymer, such as a DNA molecule, moves through the spatially periodic bands. In one preferred embodiment, an elongated polymer is labeled with one or more fluorescent labels. As the elongated polymer moves through the bands, any bound fluorescent labels will emit fluorescence in proportion to the illuminating light. The movement of the fluorescent labels will therefore result in a temporally periodic emission.

The time pattern of emission can be measured, and by knowing the spatial periodicity, i.e., the fringe spacing, the velocity of the DNA can be calculated from the measured temporal periodicity. In a related embodiment, the backbone of an elongated DNA molecule is stained with a fluorescence dye. In this embodiment, a staircase pattern of emission will be observed as successively more of the bright bands illuminate DNA. While not intending to be limited as such, the fringe pattern can be created by overlapping, in a single illumination spot, two beams 352, 354 that have been formed from a single laser and passing those beams through a microscope objective lens 356, as represented schematically in FIG. 9. The resulting fringe spacing is one half the wavelength of the illuminating light when the two beams are oriented 180 degrees to each other. At other angles, the fringe spacing is larger and can be found from $D = \lambda / 2 \sin(q)$, where D is the fringe spacing, λ is the wavelength of the illuminating laser and q is the half-angle between the two laser beams. The illumination spot preferably is a few microns in size such that a sufficient number of fringes are formed.

The optical methods described above can also be combined. For example, elongated DNA molecules labeled with both intercalating dye and sequence specific fluorescence markers are contemplated. Preferably, the intercalating dye and the fluorescence markers emit light of different and distinguishable wavelengths. In such an embodiment, signal amplitude profiles illustrated by both FIG. 3 and FIG. 5 are simultaneously determined for a DNA molecule. Any dual color configurations for measuring fluorescence from both the intercalating dye and the fluorescence markers can be used. Such configurations are described, for example, in Deniz et al., 1999, Proc. Natl. Acad. Sci. USA. 96:3670-3675; and Ha et al., 1996, Proc. Natl. Acad. Sci. USA. 93:6264-6268.

Any methods known in the art for elongating single macromolecule polymers, such as single DNA molecules, can be used in the invention. These include the methods described in

USSN 09/636,793. Single elongated polymers are then delivered to the detection region of the apparatus which is preferably located along the path of the elongated molecule to allow measurements of molecular properties of the molecule to be performed.

Center-of-Mass Velocity of Polymers

5 Various methods can be used to determine the COM velocity of an elongated polymer. Any detection methods that allow measuring the mass of the section of an elongated polymer falling within a detection zone can be used. Such detection methods can be, but are not limited to, optical detection methods and electrical detection methods.

10 In a preferred embodiment, the COM velocity is determined from measured intensity-time curves of a DNA molecule stained with intercalating dye molecules. It is well known in the art that the amount of intercalating dye molecules bound to a DNA molecule is proportional to the length of the DNA molecule. For example, intercalating dye YOYO-1 binds to DNA molecules at a dye:base pairs ratio of 1:5 (see, e.g., Larsson et al., 1994, J. Amer. Chem. Soc. 116:8459-8465). The fluorescence intensity of the intercalating dye is
15 therefore proportional to the length or mass of the DNA molecule. Thus, the center-of-mass of a DNA molecule can be determined from the measured fluorescence intensity-time curve by integrating the fluorescence intensity as a function of time according to the equation:

(1)

$$L_{com} = \frac{\int_{t_1}^{t_2} I(t) \cdot L(t) dt}{\int_{t_1}^{t_2} I(t) dt}$$

20 Where L_{com} is the location of the center-of-mass along the polymer, $I(t)$ is the fluorescence intensity measured at time t , t_1 and t_2 designate the time when the polymer enters and leaves

$$L(t) = \int_{t_1}^t v(\tau) d\tau$$

the detection zone, respectively, and $L(t)$ is the length of the polymer that has passed through the detection zone at time t . Any time before t_1 can be used as the lower bound of the integral and any time after the time t_2 can be used as the upper bound of the integral. $L(t)$ can be determined according to equation (2)

(2)

where $v(t)$ is the velocity of the polymer at time t . Thus the temporal location of the center-of-mass t_{com} of an elongated polymer can be determined by solving the equation

$$\int_{t_1}^{t_2} v(t) dt = \frac{\int_{t_1}^{t_2} I(t) \cdot \left[\int_{t_1}^t v(\tau) d\tau \right] dt}{\int_{t_1}^{t_2} I(t) dt} \quad (3)$$

5

In cases where the velocity of the polymer is constant, i.e., $v(t)=V$, the temporal location of the center-of-mass can be determined by the equation

$$T_{com} = \frac{\int_{t_1}^{t_2} I(t) \cdot t dt}{\int_{t_1}^{t_2} I(t) dt} \quad (4)$$

10 The COM velocity is then determined by dividing the distance between two detection zones of known distance with the difference in T_{com} at the two detection zones.

In another embodiment, the COM velocity is determined from the measured intensity-time curve of intrinsic stimulated fluorescence from an elongated polymer. When monomer units of a polymer can be excited by light and fluoresce at a detectable wavelength, intensity-time curves can be measured by monitoring fluorescence at the emitted wavelength. When
15 this approach is to be used, it is preferable that the fluorescent monomers are evenly distributed along the polymer. The fluorescence intensity then provides a measure of the length or mass of the polymer.

Center-to-Center Velocity of Polymers

20 Depending on the channel geometry used for delivering the polymers, a center-to-center approximation of the velocity of the polymer can be used. For example, a center-to-center velocity of an elongated DNA can be used where a constant flow profile is generated. The center-to-center velocity is generally different from the center-of-mass velocity. The

center-to-center velocity is determined by measuring the time interval for the mid point of the molecular contour of an elongated polymer, i.e., the center of the polymer contour, to travel between two detection zones separated by a known distance. The center of an elongated polymer is defined as

5

$$L(t_c) = \frac{L(t_1) + L(t_2)}{2} \quad (5)$$

where $L(t_c)$, $L(t_1)$ and $L(t_2)$ are defined by Eq. (2). In cases where the velocity of the polymer is constant, i.e., $v(t)=V$, the temporal location of the center of the polymer can be determined by the equation

10

$$t_c = \frac{t_1 + t_2}{2}$$

(6)

where t_c is the temporal location of the center of the measured signal amplitude profile, e.g., intensity-time curve, of the polymer, and t_1 and t_2 are the leading and trailing edges of the signal amplitude profile, respectively. The center-to-center velocity is then determined by dividing the distance between two detection zones with the difference in t_c at the two detection zones.

15

In some cases, e.g., when a polymer is elongated such that its mass is uniformly distributed across its length, the center-to-center velocity is the same as the center-of-mass velocity.

20

End-to-End Velocity of Polymers

The end-to-end velocity is determined using the entrance or exit of the elongated molecule through two detection zones. In some embodiments, a leading edge to leading edge velocity is determined by measuring the leading edge of the polymer passing through the two detection zones. In other embodiments, a trailing edge to trailing edge velocity is determined by measuring the trailing edge of the polymer passing through the two detection zones. End-

25

to-end velocity can be measured by any methods that are capable of detecting the ends of a polymer.

In some embodiments of the invention, the end-to-end velocity of an elongated polymer is determined from signal amplitude profiles measured at two detection zones by a method that allows measuring the mass of polymer a section in each detection zone. In a preferred embodiment, the polymer is a DNA molecule labeled along its entire length with an intercalating dye. The entrance of the DNA molecule into a detection zone is marked by an increase in the intercalator signal. The exit of the DNA molecule from the detection zone is marked by a decrease in the intercalator signal. Using the known distance of separation in the detection zones, the velocity can be determined using any combination of edges of the intensity-time curves and their respective times of entry/exit into their respective detection zones. In one embodiment, a time of entry or exit of the polymer in the detection zone is identified as the temporal location where intensity-time curve begins to deviate from the background level. In another embodiment, a time of entry or exit of the polymer is identified as the time at the half height of the leading or trailing edge of the intensity-time curve. The end-to-end velocity is then determined by dividing the distance between two detection zones with the difference in the entry or exit times at the two detection zones.

In other embodiments, the polymer is labeled at one or both ends and the label is detected to determine the end-to-end velocity. For example, the polymer can be labeled at one end with a fluorescent dye molecule. The fluorescence signals are detected at two detection zones separated by a known distance. The end-to-end velocity is determined by dividing the distance between the two detection zones by the measured time interval between the signals. Because the polymer can enter the detection zones with either its labeled end or its unlabeled end as the leading edge, it is important that the polymer travel from the first detection zone to the second detection zone without switching leading and trailing ends. This condition is normally satisfied if no obstacles are located between the two detection zones.

In a preferred embodiment, the polymer is labeled at both ends with the same dye. Two fluorescence signals are detected at each detection zone. The end-to-end velocity can be determined by the two first signals or the two second signals. Labeling of the polymer at both ends is more preferred than labeling of the polymers at one end, since the detection of two correlated signals corresponding to the labels at the leading and the trailing edges at each detection zone can be used to confirm the detection of the labeled ends of a polymer. This is especially important when loose dye molecules or other fluorescent species may be present

in the sample. In another preferred embodiment, the two ends of a polymer can be labeled with dyes having different emission spectra.

Velocity of Polymers from Rise-Time Measurements

5 The rise-time approximation allows the determination of the velocity of an elongated polymer by dividing the time required for the leading edge of the polymer to traverse the detection zone into the known length of the detection zone. Alternately, the velocity of an elongated polymer can be determined by dividing the time required for the trailing edge of the polymer to traverse the detection zone into the known length of the detection zone. In
10 contrast to COM, center-to-center and end-to-end velocities, the rise-time velocity is determined using only one detection zone.

 In some embodiments of the invention, the rise-time velocity of an elongated polymer is determined from signal amplitude profiles measured at a detection zone by a method that allows measuring the mass of polymer section in the detection zone. In a preferred
15 embodiment, the polymer is a DNA molecule labeled with intercalating dye. The entrance of the DNA molecule into a detection zone is marked by a rise in the intercalator signal. The exit of the DNA molecule from the detection zone is marked by a decrease in the intercalator signal. The intercalator signal from the DNA reaches full intensity as the front end of the DNA molecule reaches the far-edge of the detection zone. Thus, in an intensity-time curve,
20 rise-time is represented by the leading edge of the intensity-time curve. The rise-time velocity is then determined by dividing the detection zone length by the time interval in which it takes the leading end of the signal amplitude profile to reach full intensity. In one embodiment, the detection zone is defined by the spot size of a laser excitation beam from a 532 nm diode pumped solid-state laser focused to a beam waist of about 266 nm. In that
25 embodiment, the resolution of rise-time velocity is 266 nm divided by the rise-time.

Determination of the Velocity of Polymers Labeled with Markers at Known Distances on the Polymer

 The velocity of elongated polymers can also be determined by monitoring the time
30 interval between markers on the polymers at known distances. In such embodiments, because the distance between two or more markers on a polymer is known, the velocity of the polymer can be determined using a single detection zone.

In a preferred embodiment, sequence specific markers are used to label the DNA molecule. Such markers can comprise, for example, fluorescently labeled oligonucleotides of defined sequences that hybridize to and therefore “mark” complementary sequences present on a target DNA molecule. The distance between each adjacent pair of markers is preferably
5 greater than half the dimension of the detection zone so that signals from adjacent markers can be distinguished. The velocity of the polymer is then determined by dividing the known distance between a pair of markers with the time interval between detection of the signals from the two markers. Any sequence specific markers can be used in the present invention. These include, but are not limited to, fluorescently labeled sequence specific major or minor
10 groove binders and intercalators, sequence-specific DNA or peptide binding proteins, sequence specific peptide nucleic acids (PNAs), and the like.

Two or more of the methods described above can be combined to give additional information about the motion of elongated polymers. In one embodiment, COM velocity and end-to-end velocity are determined and compared for an elongated polymer. A difference
15 between COM velocity and end-to-end velocity indicates that the length of the polymer may have been changing as the polymer passed through the detection zones. In another embodiment, rise-time velocity can be used to estimate the size of the detection zone. The rise-time velocity is subtracted from a simultaneously determined COM velocity, center-to-center velocity and/or end-to-end velocity such that the contribution of the finite size of the
20 detection zone is eliminated.

Determination of Polymer Lengths

The invention also provides methods for determining lengths of single elongated polymers and for determining distances between tags along single elongated polymers. The
25 length of the polymer or the distance between two markers on a polymer, can be determined using discrete detection zones if the velocity of the polymer is known. For example, if the velocity of a polymer is $v(t)$, then the length of the polymer can be determined from the signal amplitude profile measured at a detection zone according to the equation:

$$L = \int_{t_1}^{t_2} v(t) dt$$

(7)

where L is the length of the polymer, and t_1 and t_2 are the leading and trailing edges of the signal amplitude profile.

In some embodiments of the invention, the velocity of the polymer is approximated by a time independent velocity V . In such embodiments, the length of the polymer can be
5 determined as $L = V \cdot (t_1 - t_2)$.

However, since the velocity of the polymer may be changing upon passage through the region of interest, it is preferable to provide multiple detection zones along the path of the elongated polymer so that multiple signal amplitude profiles can be obtained. In one embodiment, the spacings between each adjacent pair of detection zones are much smaller
10 than the length of the elongated polymer. In such an embodiment, overlapping signal amplitude profiles along the strand of the elongated polymer can be obtained, permitting determination of time dependent velocity $v(t)$. This will increase the accuracy of the velocity measurements and allow a more accurate measurement of the polymer length. A schematic of the output is shown in the following FIG. 4. In FIG. 4, four signal amplitude profiles of an
15 elongated polymer are shown (with arbitrary relative intensities). Any of the different types of velocities can be used in the multiple detection scheme. For instance, velocity determination can further be estimated by using a combination of leading edge velocity information, center-of-mass estimations, rise time estimations, and other information that can be obtained from the intercalator signal.

20 Single-Molecule Restriction Mapping

The methods and apparatuses of the invention can be used in analysis of polymers, such as for example, single-molecule restriction fragment length polymorphism (RFLP). In one embodiment, a suitable restriction enzyme or enzymes are used to produce restriction fragments. Any restriction enzymes can be used in conjunction with the invention.
25 The recognition sequences and reaction conditions of many restriction enzymes are well known to one skilled in the art. Depending on the sequences of the DNA molecules that are to be analyzed, suitable restriction enzymes can be selected. A sample containing fragments from restriction digestion is then labeled. For example, labeling may comprise staining the sample with an intercalating dye, such as YOYO-1. The labeled fragments are then
30 elongated and detected by any method of the invention. Two or more detection zones are located in the region where the lengths of single restriction fragments are measured.

In another embodiment, modified restriction enzymes that recognize and bind to their recognition sequences but do not cleave the substrate DNA molecule can be used to label a

DNA molecule. Alternately, a restriction enzyme can be used in conjunction with a buffer, such as a solution of low concentration Mg^{2++} , such that the restriction enzymes recognize and bind to their recognition sequences but do not cleave the substrate DNA molecule. In each of these cases, the restriction enzymes are labeled using, for example, fluorescent labels.

5 DNA molecules labeled in this manner are then elongated and distances between labels along the DNA molecules are measured. Of course, other sequence specific labels can be used in place of or in conjunction with labeled non-cleaving restriction enzymes. Such sequence specific labels include but are not limited to sequence specific major or minor groove binders and intercalators or sequence specific PNAs, and the like.

10 A combination of cleaving and non-cleaving restriction enzymes can also be used. In one embodiment, restriction enzymes that cleave a substrate can be used to generate restriction fragments. The resulting restriction fragments are stained with an appropriate intercalating dye. Non-cleaving restriction enzymes labeled with a fluorescence dye of a wavelength different and distinguishable from that of the intercalating dye are then used to
15 label internal restriction sites in these restriction fragments. Simultaneous detection of both wavelengths allows determination of lengths of restriction fragments and identification of internal sequences. Other sequence specific labels can be used in place of or in conjunction with labeled non-cleaving restriction enzymes. As above, such sequence specific labels include but are not limited to sequence specific major or minor groove binders and
20 intercalators or sequence specific PNAs, etc.

EXAMPLE

Stretching of Phage λ -DNA

FIGS. 11A-11C, 12 and 13 were derived using a dual laser spot arrangement as
25 illustrated in FIG. 10. The DNA was driven through a quartz chip having an etched nanochannel of the following design: A nanochannel was formed in a 300 nm quartz chip using electron beam lithography. The process employed a series of lithographic steps known in the art, including a) coating the quartz wafer with a resist, b) exposing a pattern on the coated chip to an electron beam, c) stripping the exposed resist, d) etching using reactive ion
30 etch to give straight wall profiles, and e) removing the remaining resist.

The detection zones were laser spots formed by an argon ion laser running at 488 nm and delivering about 2 mW of laser power to each of the spots. The spots were diffraction limited, with a spot size of approximately 0.5 microns in diameter. The laser spots were

spaced 15 mm apart and located within the nanochannel in the path of the DNA molecules. A carrier liquid containing labeled DNA was flowed through the nanochannel in a manner such that laminar flow lines allowed the DNA to be delivered through the two laser spot detection zones. The DNA molecules comprised λ -DNA that was intercalated with YOYO-1 at a final concentration of one YOYO-1 molecule/10 base-pairs. The carrier solution contained 100 mM DTT in a 1x TBE solution. A fluid drive, delivering 50 psi, was used to drive the DNA through the channel. The chip was been sealed using known chemical activation sealing techniques. The surfaces to be bonded (the quartz chip and a quartz coverslip) were activated using an ammonium hydroxide:peroxide:water mixture, heated to 70 °C. The coverslip and the quartz chip were then pressed together under water and the clamped package was baked to evaporate any solvent and cause chemical bonding of the two surfaces. The DNA sample mixture was then introduced into the chip by capillary forces. The signals measured from the two laser excitation spots, arranged 15 microns apart, were collected through a Nikon 1.4 NA 100x oil immersion objective. The resultant fluorescent signals were filtered and detected by two avalanche photodiodes. The signals from the two intercalator signals were then collected through a data capture board, processed through an A/D converter, and stored on a computer. The computer allowed processing of the data and calculation of the lengths of the molecules. A constant velocity estimation was used to derive the velocity of the molecules moving through the system. The velocity was a COM to COM velocity estimation/calculation. The average length of the lambda molecules was measured to be around 18 microns, corresponding to an approximate 20% increase in the length of the DNA molecule from the intercalator staining. The data was processed using an algorithm written in MATLAB data processing language.

FIGS. 11A-11C show exemplary signal amplitude profiles measured at two detection zones.

FIG. 12 shows a velocity histogram based upon the center-of-mass velocity of the detected λ -DNA molecules.

FIG. 13 shows a length histogram of the of λ -DNA molecules.

What is claimed is:

CLAIMS

1. A method for determining velocity of an elongated polymer, the method comprising:

- 5 a) defining at least one detectable region on the polymer;
- b) causing relative movement of the elongated polymer through a plurality of linearly sequential detection zones, each separated a predetermined distance, to cause interactions between the detectable region and the detection zones, thereby producing a plurality of signal amplitude profiles, each signal amplitude profile being produced at a different detection zone and comprising
- 10 data acquired before, during and after each interaction;
- c) measuring each of the signal amplitude profiles in a time-correlated manner; and
- d) analyzing the time-correlated measurements to determine the velocity of the polymer.

15

2. The method of claim 1, wherein the plurality of linearly sequential detection zones comprises a first detection zone and a second detection zone.

3. The method of claim 2, wherein a first signal amplitude profile is

20 measured at the first detection zone and a second signal amplitude profile is measured at the second detection zone.

4. The method of claim 3, wherein each signal amplitude profile has a center-of-mass, a contour center, a leading edge, and a leading-edge-rise-time.

25

5. The method of claim 4, wherein the velocity of the elongated polymer is a center-of-mass velocity determined by the steps of:

a) determining the time interval between detection of the center-of-mass in the first signal amplitude profile and the center-of-mass in the second signal amplitude profile; and

b) dividing the predetermined distance between the detection zones by the time interval.

6. The method of claim 4, wherein the velocity of the elongated polymer is a center-to-center velocity determined by the steps of:

a) determining the time interval between detection of the contour center of the first signal amplitude profile and the contour center of the second signal amplitude profile; and

b) dividing the predetermined distance between the detection zones by the time interval.

7. The method of claim 4, wherein the velocity of the elongated polymer is an edge-to-edge velocity determined by the steps of:

a) determining the time interval between detection of the leading edge of the first signal amplitude profile and the leading edge of the second signal amplitude profile; and

b) dividing the predetermined distance between the detection zones by the time interval.

8. The method of claim 4, wherein the velocity of the elongated polymer is a rise time velocity determined by the steps of:

a) determining the time interval between detection of the leading-edge-rise-time in the first signal amplitude profile and the leading-edge-rise-time in the second signal amplitude profile; and

b) dividing the predetermined distance between the detection zones by the time interval.

5 9. The method of claim 1 wherein the detectable region is intrinsically detectable.

 10. The method of claim 1 wherein the detectable region is extrinsically detectable.

10 11. The method of claim 1 wherein the detectable region is detected by measurement of a physical quantity selected from a group comprising of electromagnetic radiation, electrical conductance, thermal conductance, and radioactivity.

15 12. The method of claim 1 wherein the detectable region is detected by direct or indirect measurement of fluorescent radiation.

 13. A method for determining the length of an elongated polymer, the method comprising:

20 a) defining a detectable region along the entire length of the polymer;

 b) causing relative movement of the elongated polymer through first and second detection zones, the zones being linearly spaced apart by a predetermined distance;

25 c) measuring a time interval between detection of the elongated polymer at the first detection zone and detection of the elongated polymer at the second detection zone;

d) dividing the predetermined distance between the first and second detection zones by the time interval of step c) to determine the velocity of the polymer;

e) measuring, at one of the detection zones, a time interval during which the polymer is detected; and

f) multiplying the velocity of step d) by the time interval of step e) to determine the length of the elongated polymer.

14. A method for determining the distance between first and second landmarks on an elongated polymer, the method comprising:

a) providing first and second landmarks on an elongated polymer;

b) causing relative movement of the elongated polymer through first and second detection zones, the zones being linearly spaced apart by a predetermined distance, to cause detection of the first and second landmarks at the first detection zone and detection of the first and second landmarks at the second detection zone;

c) measuring the time interval between detection of one landmark at the first detection zone and detection of that same landmark at the second detection zone;

d) dividing the predetermined distance between the first and second detection zones by the time interval of step c) to determine the velocity of the polymer;

e) measuring the time interval between detection of the first landmark at one detection zone and detection of the second landmark at that same detection zone; and

f) multiplying the velocity of step d) by the time interval of step e) to determine the distance between the first and second landmarks.

15. The method of claim 1, wherein the elongated polymer comprises an elongated DNA molecule.

16. The method of claim 13, wherein the elongated polymer comprises an elongated DNA molecule.

17. The method of claim 14, wherein the elongated polymer comprises an elongated DNA molecule.

18. The method of claim 1, wherein the elongated polymer includes fluorescent labels, and further wherein the measurements are measurements of fluorescence intensity.

19. The method of claim 13, wherein the elongated polymer includes fluorescent labels, and further wherein detection of the elongated polymer comprises detection of fluorescent energy.

20. The method of claim 14, wherein each of the landmarks comprises a fluorescent label, and further wherein detection of the landmarks comprises detection of fluorescent energy.

21. The method of claim 20, wherein the first landmark is labeled with a first fluorescent tag and the second landmark is labeled with a second fluorescent tag, and further wherein the first and second fluorescent tags emit fluorescent energy at different and distinguishable wavelengths.

22. The method of claim 1, wherein the detection zones are positioned along an elongation channel through which the polymer is caused to travel.

23. The method of claim 13, wherein the detection zones are positioned along an elongation channel through which the polymer is caused to travel.

5 24. The method of claim 14, wherein the detection zones are positioned along an elongation channel through which the polymer is caused to travel.

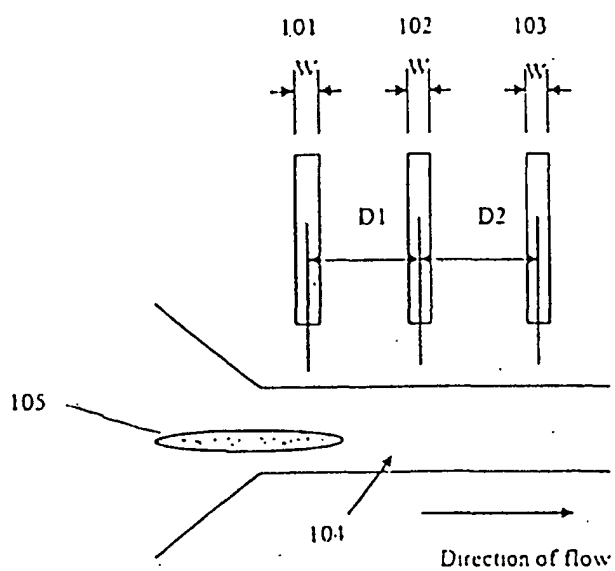


Fig. 1

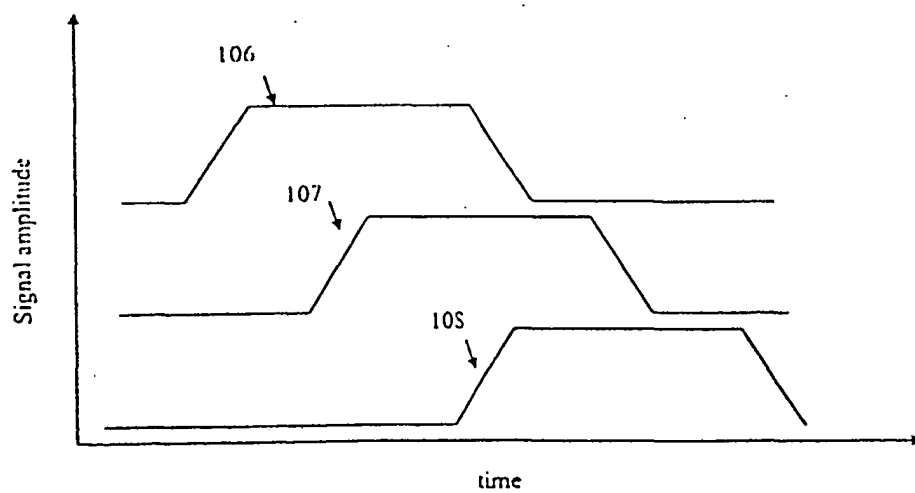


Fig. 2

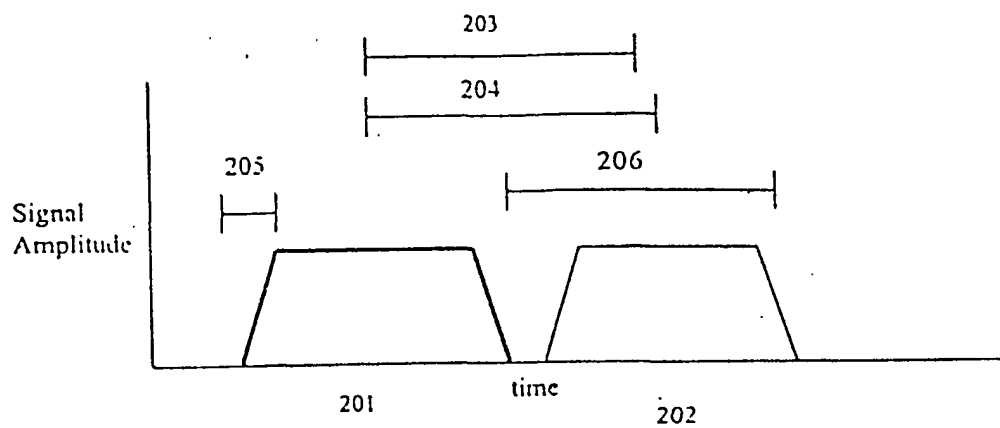


Fig. 3

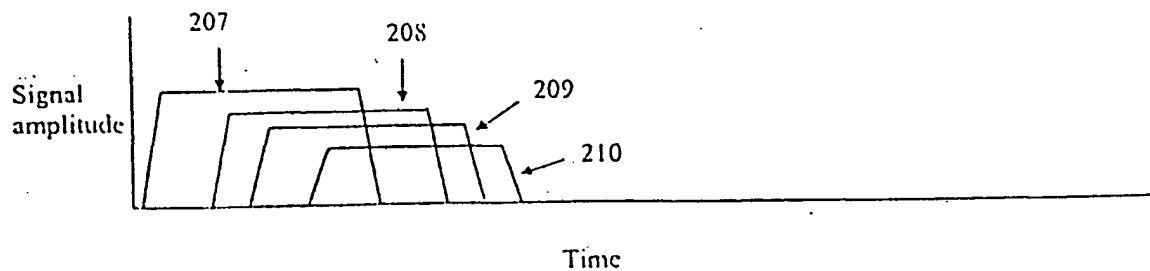


Fig. 4

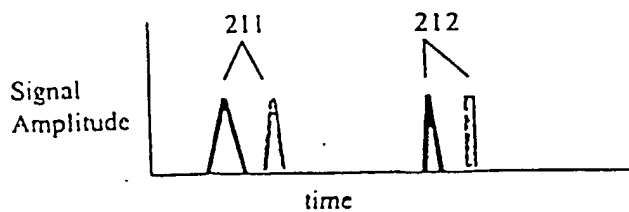
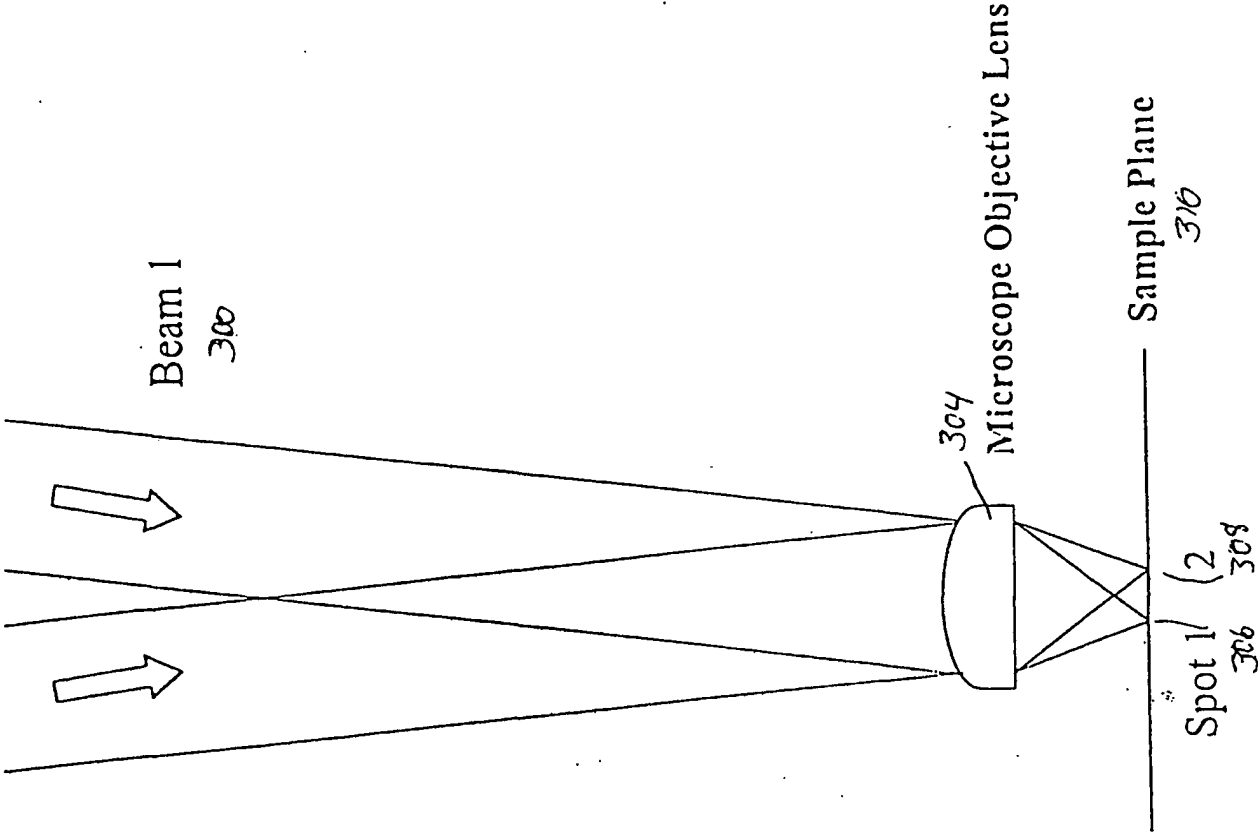


Fig. 5



Optical
Arrangement for
Multiple Spot
Formation

FIG. 6

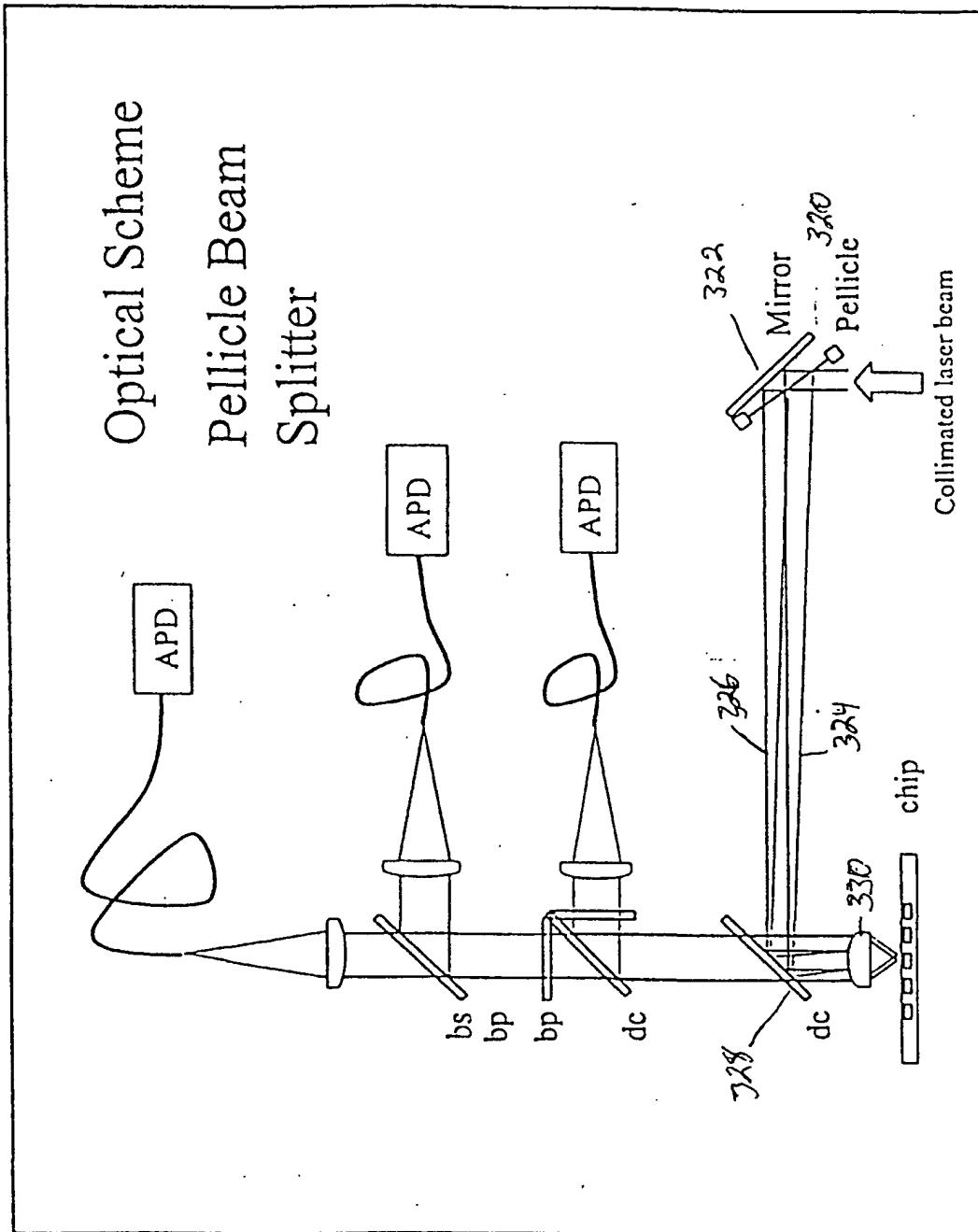


FIG. 7

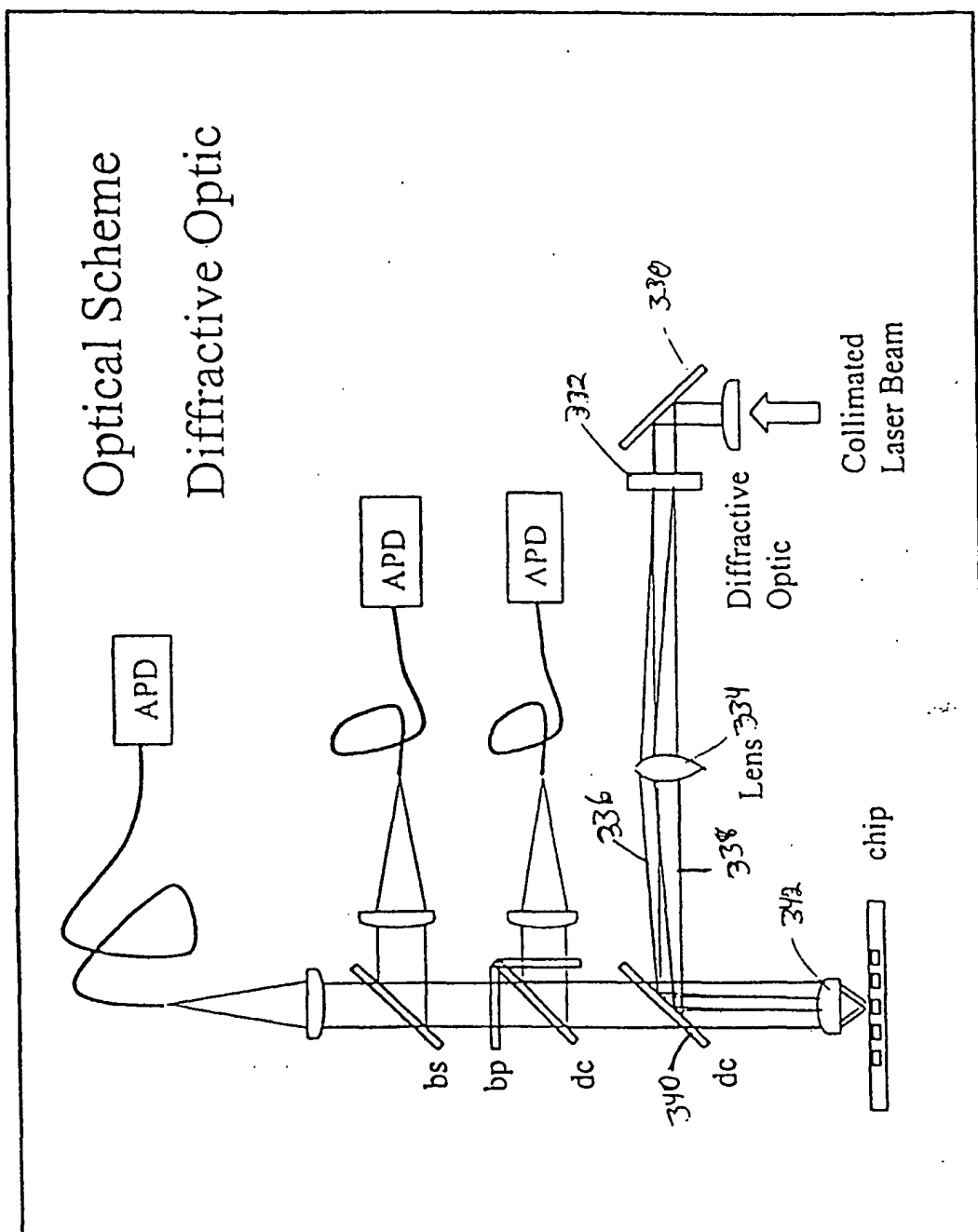


FIG. 8

Optical Arrangement for Single Spot with Interference Fringes

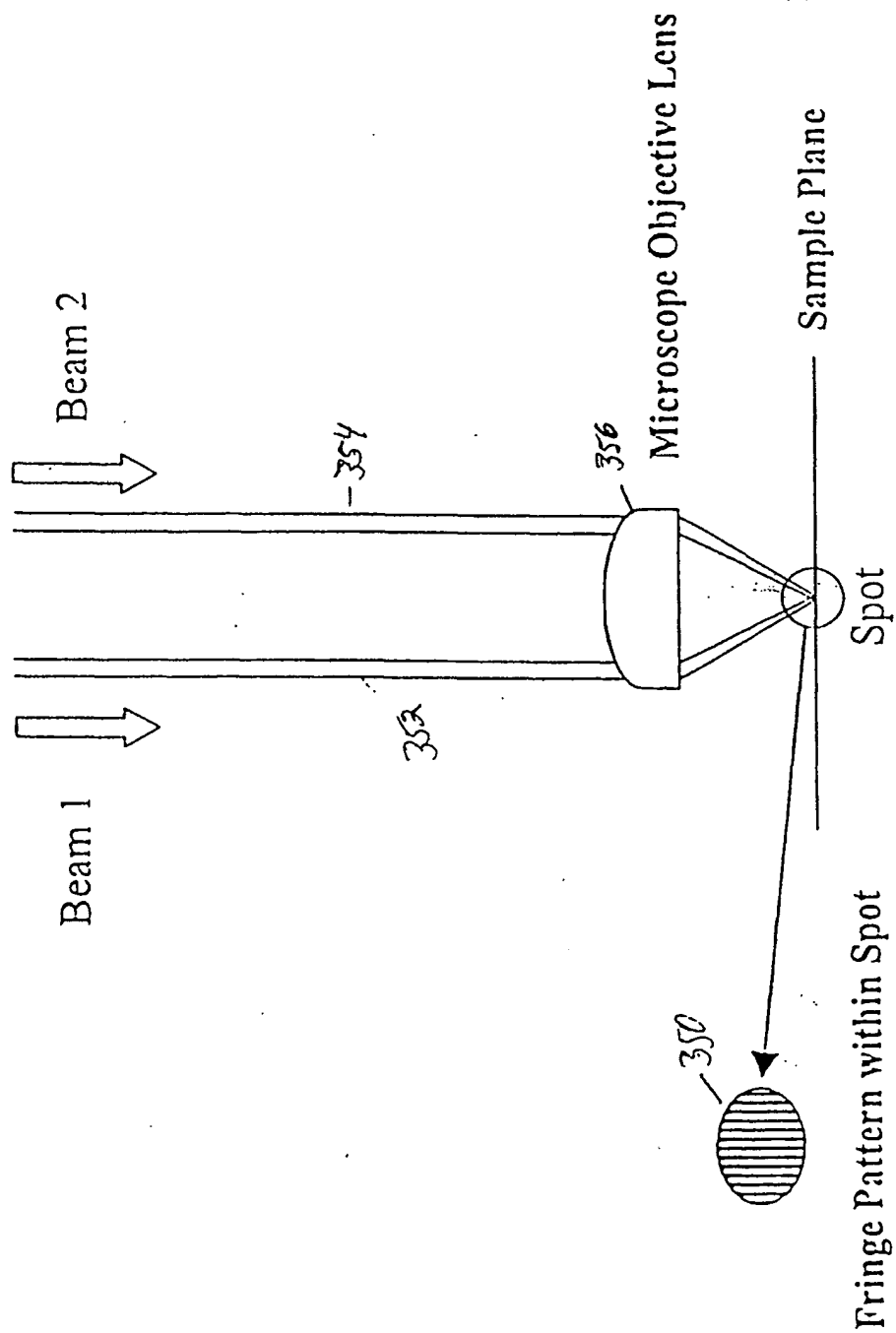


FIG. 9

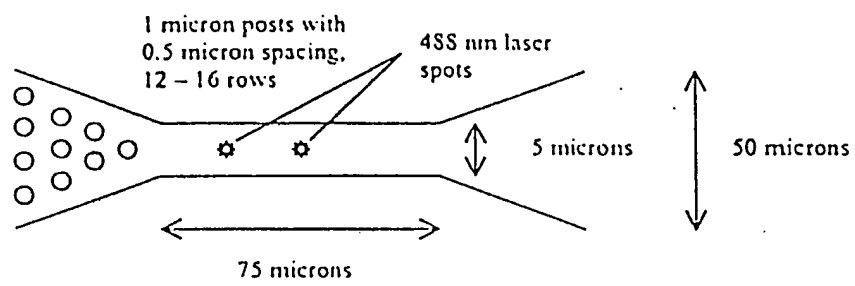


FIG. 10

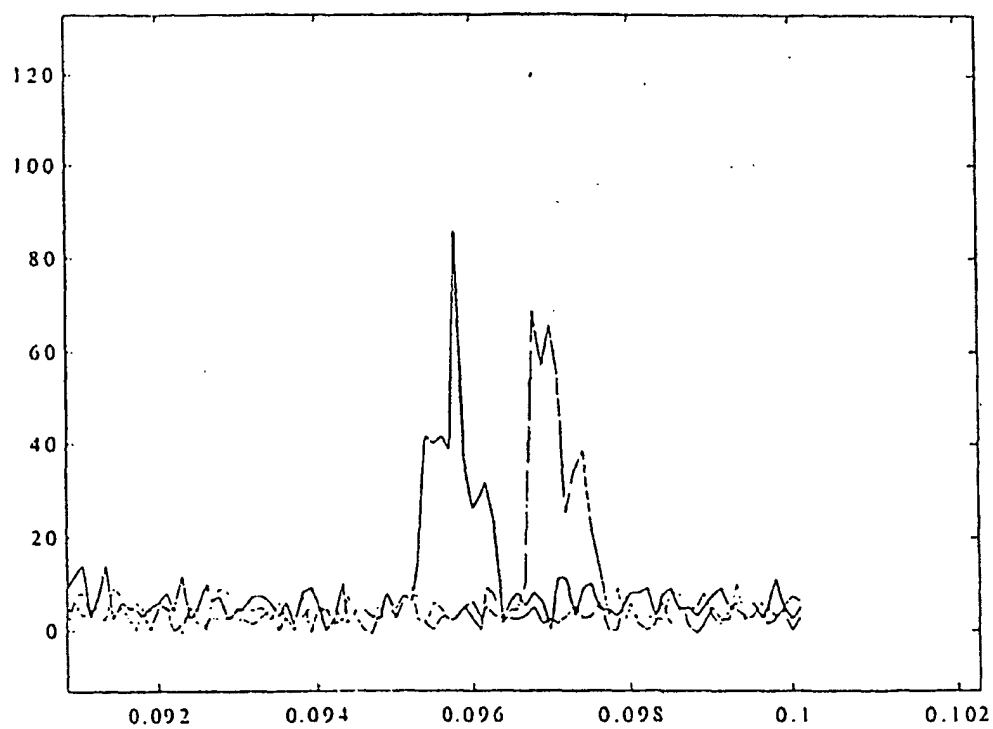


Fig. 11A

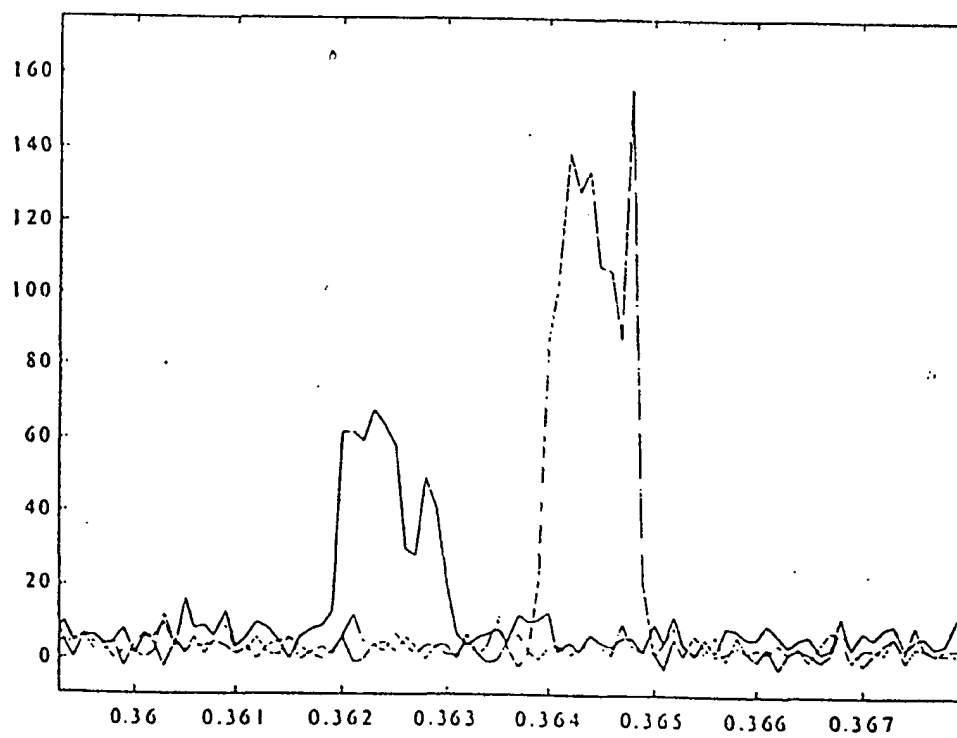


Fig. 11B

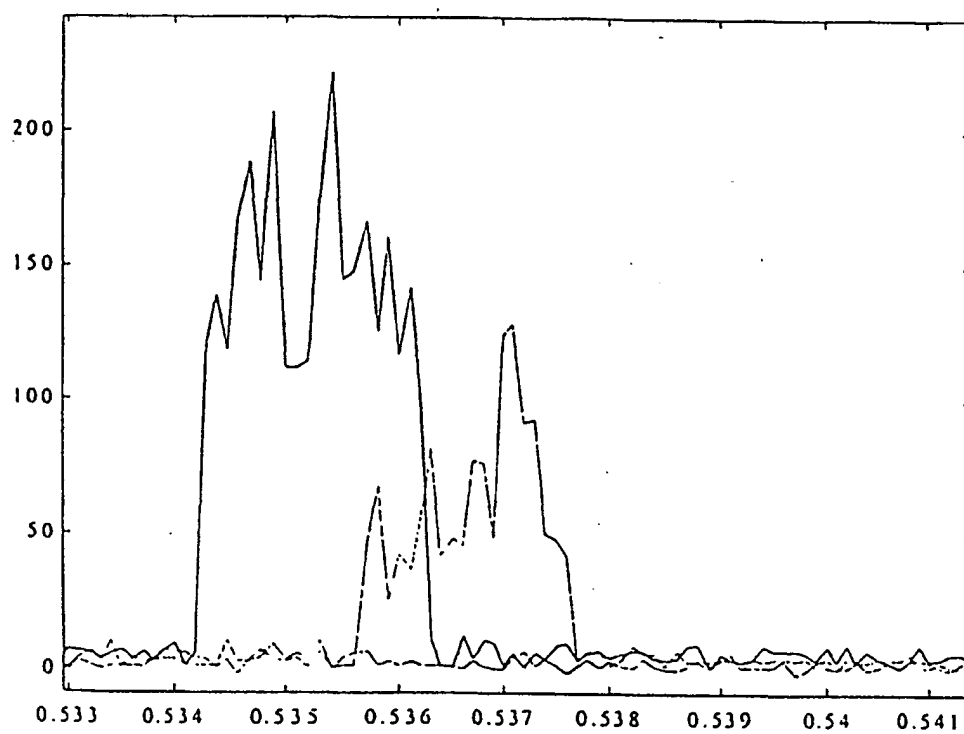


Fig. 11C

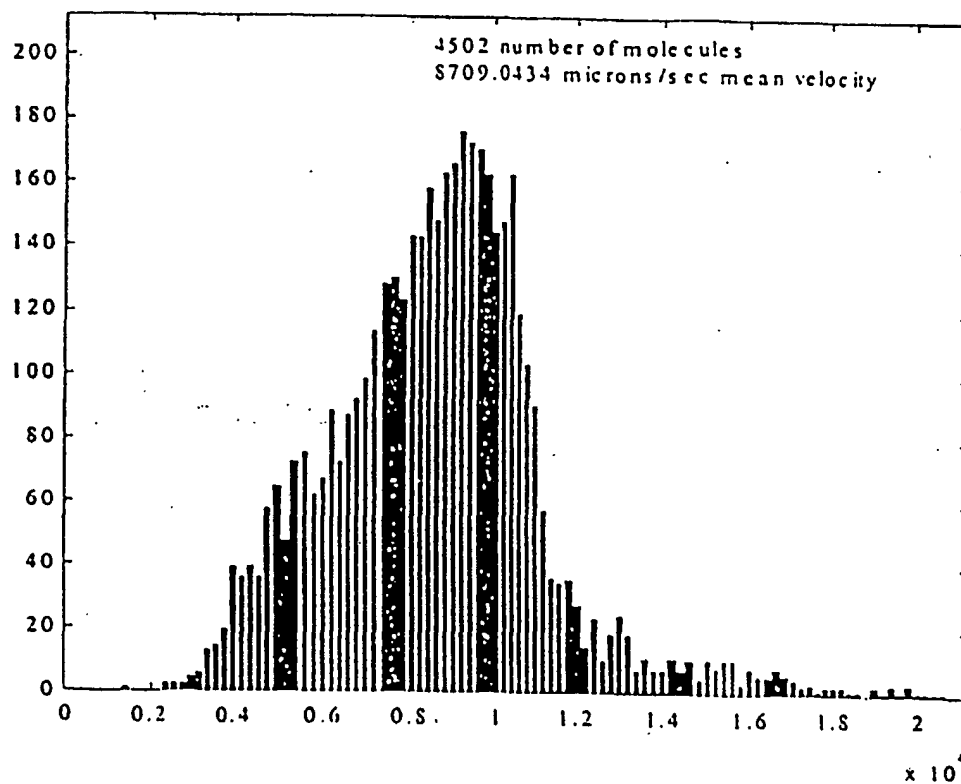


Fig. 12

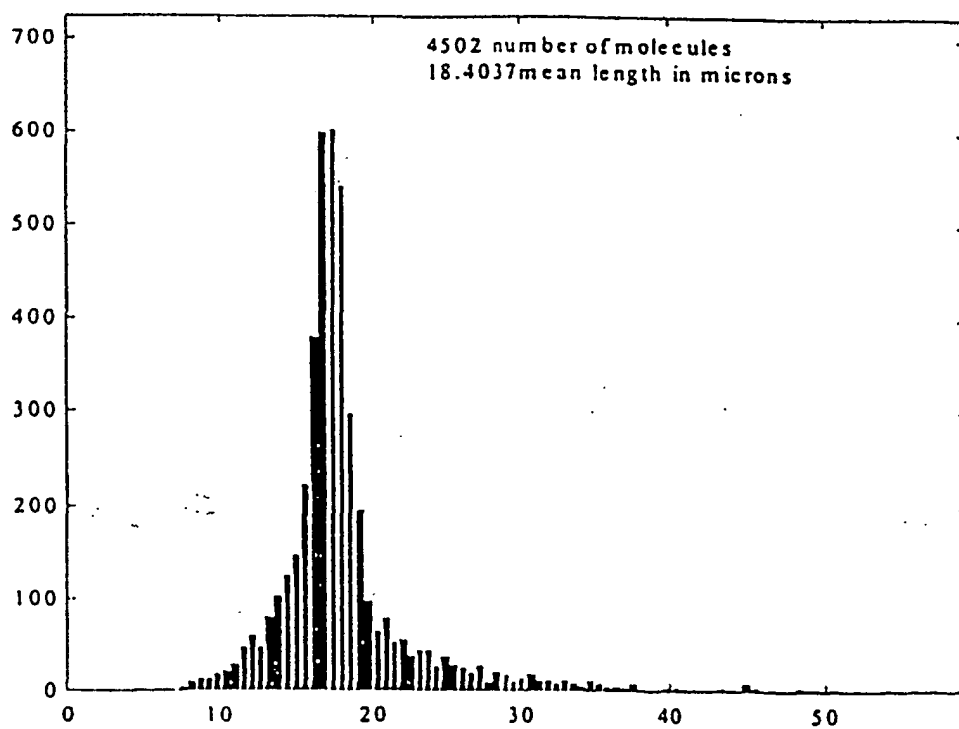


Fig. 13

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/17124

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G01N 21/64

US CL : 436/94, 164, 172; 435/6

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 436/94, 164, 172; 435/6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98/10097 (PIONEER HI-BRED INTERNATIONAL, INC.) 12 March 1998 (12.03.1998), page 7, lines 4-17.	1-12, 15, 18, 22
A	US 5,356,776 A (KAMBARA et al) 18 October 1994 (18.10.1994), column 3, lines 5-20.	13, 16, 19, 23
A	US 5,707,797 A (WINDLE) 13 January 1998 (13.01.1998), column 3, lines 45-60.	14, 17, 20-21, 24

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"

document member of the same patent family

Date of the actual completion of the international search

23 August 2002 (23.08.2002)

Date of mailing of the international search report

19 SEP 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks

Box PCT

Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Jan M. Ludlow

Telephone No. (703) 308-0661

DEBORAH THOMAS

LEGAL SPECIALIST

INTERNATIONAL SEARCH REPORT

PCT/US02/17124

Continuation of B. FIELDS SEARCHED Item 3:

EAST search terms: stretch\$ or eleongat\$; measur\$ or determin\$ or sensing or sensor\$ or analyz\$; speed or velocit\$ or length or distance; polymer\$ or DNA or RNA or protein or peptide or polypeptide or polysaccharide or oligopeptide or oligosaccharide or oligonucleotide or oligonucleoside or polynucleotide or polynucleoside